



Řešení pro ukládání dat

Zaostřeno na SDS i na bezpečnost

Redakce BusinessIT a partneři

Řešení pro ukládání dat: Zaostřeno na SDS i na bezpečnost

BusinessIT.cz

Edice: BusinessIT ebooks

Autoři: Redakce BusinessIT.cz

Copyright © Bispiral, s.r.o., 2014

Vydáno v roce 2014 v Bispiral, s.r.o.

Názvy použité v této knize mohou být ochrannými známkami příslušných vlastníků.

web: www.BusinessIT.cz

Explozivního růstu objemu vytvářených dat jsme si všichni vědomi, stejně jako základních faktorů, které k němu vedou. Jaký vývoj ale lze čekat v blízké budoucnosti? A jaké významné trendy lze vysledovat v oblasti ukládání dat? Právě na tato témata se zaměříme v této e-knize. A nevyhneme se ani otázce, co s citlivými daty na konci jejich životního

cyklu.

Podle studie EMC Digital Universe vycházející z údajů IDC mezi lety 2013 a 2020 vzroste objem vygenerovaných dat 10× – z 4,4 bilionů gigabajtů (zettabytes) na 44 bilionů gigabajtů (zettabytes).

Jaké budou hlavní příčiny? A jak se vyvíjí trh se systémy pro ukládání dat? Těmto tématům se věnujeme v prvním příspěvku této e-knihy.

Ve druhé kapitole jsme pak zaostřili na významný trend posunu směrem k softwarově definovaným systémům pro ukládání dat - na jejich principy a praktické přínosy. A ve třetí kapitole se pak zabýváme otázkou, jak bezpečně smazat data ze systémů pro ukládání dat. Někdy to může být větší problém, než si je leckdo z nás ochoten připustit.

Redakce BusinessIT.cz

Partnerem této eknihy je:



Trh storage a 44 zettabajtů dat na dohled

Informace o neustálém zrychlování růstu objemu vytvořených dat jsou evergreenem, ale jak rychlý ten růst ve skutečnosti je? A jaký vývoj lze čekat v blízké budoucnosti? Analytici se shodují, že se objem dat ročně vygenerovaných na celém světě každých 12 měsíců zdvojnásobuje a jejich růst rozhodně nepoleví – mimo jiné i v souvislosti s nástupem takzvaného internetu věcí. Roste ale stejně rychle i objem instalovaných úložných systémů?

Studie EMC Digital Universe vycházející z údajů IDC tvrdí, že mezi lety 2013 a 2020 vzroste objem vygenerovaných dat 10× – z 4,4 bilionů gigabajtů (zettabytes) na 44 bilionů gigabajtů (zettabytes). Podle analytiků IDC se pak počet drobných zařízení (mimo běžná PC, tablety, chytré telefony apod.), která mají potenciál být připojena k internetu, blíží 200 miliardám, přičemž 7 % (14 miliard) je již k této síti připojených a komunikuje přes ni.

Objem dat ze zmíněných připojených zařízení představuje 2 % všech dat, která dnes na světě vznikají, a IDC odhaduje, že do roku 2020 počet

připojených zařízení vzroste na 32 miliard a budou generovat 10 % celosvětových dat. Těmito zařízeními přitom budou podle analytiků IDC nejrůznější věci – od inteligentních automobilů přes hračky a vybavení domácnosti až po průmyslová zařízení nebo třeba psí obojky.

Trh s úložnými zařízeními

Pro uložení vytvořených dat jsou pochopitelně nezbytné odpovídající úložné systémy. Podívejme se tedy, jak se daří těm. Podle analytiků společnosti Gartner v loňském roce vzrostl obrat na trhu s externími diskovými úložnými zařízeními (ECB, External Controller-Based Disk Storage) o 1,4 % na 22,5 miliardy amerických dolarů. Za nízkým růstem, nejnižším od poklesu v roce 2009, vidí analytici stále nejistou makroekonomickou situaci především v Severní Americe, EMEA a regionu Asie/Tichomoří. Ještě horší situace byla pak podle analytiků IDC situace ve střední a východní Evropě s meziročním poklesem obratu o 1 % (2,6 % v posledním čtvrtletí). Jedním dechem je ovšem třeba dodat, že díky

stálému poklesu ceny za gigabajt uložených dat to i tak znamená mírný nárůst kapacity prodaných zařízení. Hlavní trojice výrobců, EMC, HP a IBM tu přitom mají cca 75 % trhu, přičemž vede EMC, podle analytiků především díky řešením Isilon, HP se pak daří s produkty 3PAR StoreServ a IBM s produkty v tzv. entry level segmentu.

Ačkoli v loňském roce zaznamenal trh s ECB jen velmi mírný růst, ve čtvrtém čtvrtletí se konečně blýsklo na lepší časy při meziročním růstu o 5 %. Meziroční růst obrátu zaznamenalo podle analytiků Gartneru více firem, ale jen EMC a HP porazily trh; pomyslnou první příčku žebříčku pak obsadilo HP, a to zřejmě díky kladnému přijetí platform 3PAR StoreServ a StoreOnce. EMC se pak dařilo především s inovovanou nabídkou produktů VNX a Data Domain. Ztráty podílu na trhu pak podle stejného zdroje naopak zaznamenaly Dell, IBM a Oracle.

Kde vzniká nejvíce dat

Vraťme se ale ještě na chvíli k výše citované studii

EMC: Jestliže v současnosti vzniká 60 % dat v digitálním světě na takzvaných rozvinutých trzích, jako jsou Německo, Japonsko či Spojené státy, do roku 2020 se tento podíl výrazně změní, a to ve prospěch rozvíjejících se trhů - především Brazílie, Číny, Indie, Mexika a Ruska, které budou produkovat většinu dat.

Pokud jde o ukládání dat do cloudu, jejich podíl se prý během 7 let cca zdvojnásobí: V roce 2013 méně než 20 % dat v digitálním světě nějakým způsobem prošlo cloudem, do roku 2020 jich prý bude 40 %. Aby měla data pro byznys co nejvyšší přínos, měla by být snadno dostupná a být updatována v reálném čase. Jen část dat je samozřejmě významná, ta je třeba identifikovat. Firmy pak musejí definovat svou politiku ohledně dat, zajistit příslušnou hardwarovou infrastrukturu (vlastní nebo formou služby) i příslušný software pro jejich efektivní zpracování. A potřebují samozřejmě příslušné know-how.

Studie EMC rovněž upozorňuje, že růst objemu dat je rychlejší než růst kapacity úložišť, jinými slovy, že veškerá volná (tedy nevyužitá) kapacita všech úložišť na světě roste pomaleji než objem generovaných dat. V roce 2013 by volná kapacita úložišť dokázala

pojmout pouze 33 % dat digitálního světa. Do roku 2020 to prý bude méně než 15 %. Nicméně většina vznikajících dat je jen přechodná (například videostreamy ze služeb Netflix či Hulu, herní interakce na Xbox ONE, digitální televize atd.) a nevyžaduje tedy dlouhodobé ukládání.

Další vývoj trhu

Analytici IDC upozorňují, že ačkoli objem generovaných dat stále roste, vývoj na trhu ovlivňuje také řada dalších faktorů – kromě již zmiňované makroekonomické situace třeba také politická nestabilita v některých částech Evropy – a v souvislosti s ekonomicky hubenějšími uplynulými roky také sklon pečlivěji plánovat potřebnou úložnou kapacitu. Právě poslední zmiňovaný faktor znamená, že se nyní snižuje objem nevyužité kapacity úložných systémů – a jakmile rezervy poklesnou pod přijatelnou úroveň, kterou ovšem nelze objektivně stanovit, investice do nové kapacity opět zaznamenají výraznější růst.

Horký trend: Softwarově definované systémy pro ukládání dat

Jedním z aktuálních IT trendů je požadavek na směřování k vysoké flexibilitě IT systémů, a to prostřednictvím softwarově definované infrastruktury. Tento trend se nevyhýbá ani oblasti systémů pro ukládání dat, kde nachází uplatnění mimo jiné při realizaci náročných IT systémů zaměřených na práci s dalším aktuálním fenoménem – big data. Výhody ale přináší nejen tam.

Všudypřítomnost mobilních zařízení, ale třeba také předpokládaný trend nástupu internetu věcí, budou nepochybně znamenat jak nové příležitosti, tak růst objemu uložených dat. Aby tato data bylo možno efektivně využít, je třeba je ovšem nejen bezpečně uložit, ale také mít k nim dostatečně rychlý přístup, a to ze všech systémů, kde jsou potřeba. Jednou z cest, která tomu má napomoci, jsou právě softwarově definované systémy pro ukládání dat.

Softwarově definované systémy pro ukládání dat v kostce

Zjednodušeně lze říci, že softwarově definované systémy pro ukládání dat (Software-Defined Storage, SDS) označuje storage, kde je principiálně oddělen hardware pro ukládání dat a software vyšší úrovně, který ho spravuje. Stejně jako u dalších softwarově definovaných prvků IT infrastruktury se pak v optimálním případě všechny systémy využívající storage obracejí na jeden software, který zajišťuje optimální využití úložných kapacit. Tento software realizuje řízení na základě stanovených pravidel, a to včetně procesů pro deduplikaci, thin provisioning, replikaci nebo zálohování. Přitom je v optimálním případě schopen učení a tedy průběžné optimalizace všech procesů.

Další prvky IT infrastruktury, které využívají systémy pro ukládání dat, jsou odstíněny od fyzické reprezentace storage - je jim virtuálně poskytována v té podobě, jakou vyžadují, bez ohledu na skutečnou fyzickou podstatu úložných systémů. Vzhledem k tomu, že má software řídicí storage (v optimálním

případě) k dispozici všechnu fyzickou infrastrukturu pro ukládání dat organizace, může na základě daných pravidel – a znalosti aplikací, pro které poskytuje služby - zajistit její efektivní využití i příslušnou úroveň zabezpečení ukládaných dat. Systémy SDS jsou navrženy tak, aby byly schopny spravovat heterogenní hardware, a to v podobě běžně dostupných (komoditních) úložných systémů. Další jejich podstatnou vlastností je to, že jsou schopny efektivně pracovat s úložnými systémy využívajícími různých principů (klasické disky, flash a v optimálním případě i externí cloud, tedy z hlediska infrastruktury storage ve formě služby) a přitom nabízejí kompletní spektrum úložných služeb. Stejně jako v dalších oblastech IT infrastruktury i u SDS jsou preferována řešení postavená na otevřených standardech.

SDS v praxi

V reálném světě najdou SDS, tedy softwarově definované systémy pro ukládání dat, uplatnění v řadě různých scénářů – a jejich počet se bude

nepochybně stále zvětšovat jak s růstem objemů analyzovaných dat, tak s využíváním cloudů.

Představme si například celosvětově působící obchodní firmu, která v rámci svých obchodů generuje velké objemy dat na několika kontinentech. Současně zpracovává i množství externích dat, například o konkurenci (nabídce, cenách apod.) na lokálních trzích, data ze sociálních sítí apod. Pro optimální fungování je třeba mít přístup ke všem těmto datům z jednoho místa, optimálně v reálném čase.

System SDS odstíní jednotlivé aplikace od fyzických úložišť dat, přičemž zajišťuje jejich ukládání tak, aby byly (na základě zvolených pravidel) optimalizovány příslušné parametry (rychlost, zabezpečení, využití zdrojů) s ohledem na dostupnou infrastrukturu. Pokud dojde při ukládání dat k potížím – například se zpomalí přenos dat do cloudu, SDS zvolí náhradní řešení tak, aby byly splněny stanovené podmínky. Ani aplikace, jež data produkují, ani ty, které je analyzují, se nemusejí zabývat tím, kde a jak jsou data fyzicky uložena (a komplexností tohoto systému, který bude mnohdy sestávat z distribuovaných vlastních systémů i externího cloudu) – a pokud

dojde k problému, jako ve výše zmíněném případě, SDS zajistí, že aplikace, jež data vyžaduje, je dostane z příslušného systému bez toho, aby věděla o zmiňovaných potížích.

Lze očekávat, že v následujících letech se SDS stanou jedním ze zásadních prvků datových center. Pokud bude jejich implementace provedena správně, povede to k vyšší spolehlivosti, k efektivnějšímu využívání infrastruktury i ke snížení nákladů na její správu.

Citlivá data: Mažete je opravdu bezpečně?

O bezpečném mazání dat už toho bylo napsáno hodně, ale možná se dosud málo zdůrazňovalo jedno: Pokud běžní uživatelé mají problém s tím, aby jejich data skutečně spolehlivě zmizela ze světa, co teprve správci IT? Skutečně ve své organizaci mažete spolehlivě všechna citlivá data, která se mají stát minulostí – i ze všech záloh a cloudů?

Začněme právě u záloh. Jakkoli v profesionálním firemním prostředí zpravidla není třeba upozorňovat na jejich nezbytnost, s jejich vlivem na proces mazání dat už je to horší. Ne každý si totiž uvědomuje, že i když data smaže z produkčních serverů, stále ještě mohou existovat na záložních médiích.

V optimálním případě existuje v organizaci předpis, který jasně definuje, jak s daty zacházet – kdy a jak se provádějí zálohy, kde se uchovávají (ať už jde o technologii nebo lokalitu), ale také to, kdy a jak se zálohy mažou. Přitom třeba u mazání magnetických pásek je možno použít několik způsobů likvidace dat – od zdlouhavého přepisování po použití speciálních zařízení. Ta pro magnetické pásky, ale třeba i pro pevné disky nabízí třeba společnost VS Security Products (mj. s certifikací NSA).

Do hry vstupuje virtualizace

Nejde ale pochopitelně jen o problém mazání nebo ničení fyzických médií: V okamžiku, kdy vstoupí do hry virtualizace ukládání dat, může se stát situace ještě komplikovanější: Virtualizace totiž odstíjí

aplikace od fyzického úložiště, takže smazat správným způsobem ten správný disk - skutečně, ne jen zdánlivě - je najednou o poznání složitější a vyžaduje to využití odpovídajících postupů, případně nástrojů.

Jestliže u nevirtualizovaného souborového serveru mnohdy pro vyžadovanou úroveň bezpečnosti postačí prosté smazání a fyzické přepsání nevyužitého prostoru disků náhodnými daty (případně opakované), u virtualizovaného souborového serveru je třeba počítat s hypervisorem – a s pomocnými daty, která si zapisuje – například o prováděných změnách.

U rozsáhlejších systémů pro ukládání dat je třeba vzít v úvahu všechny možné repliky ukládaných dat – na různých discích i dalších médiích. Je rovněž třeba vzít v úvahu, že inteligentní systémy správy dat při práci s daty tato na základě nejrůznějších pravidel přesouvají mezi systémy – a kopie dat tak mohou být na více systémech, než byste v první chvíli čekali; ale ano, tento přístup už trochu zavání paranoiou – jak obtížné by bylo taková data najít a obnovit? Míra opatrnosti záleží na míře citlivosti mazaných dat.

Efektivní řešení

Na trhu existuje několik firem, které nabízejí speciální služby v oblasti spolehlivého mazání dat – od analýzy prostředí a prostředků, ze kterých je data nutno smazat, až po případnou fyzickou likvidaci nosičů dat. Za všechny jmenujme Kroll Ontrack nebo Sims Recycling.

Nutno ovšem dodat, že zpravidla není třeba jít až tak daleko – vždy, jak už bylo zmíněno, záleží na citlivosti mazaných dat (a tedy i na předpokládané míře úsilí, které by potenciální zájemce vynaložil na jejich získání).

Jako (zpravidla) vždy ovšem platí, že pečlivé plánování je lepší, než hledání řešení za pochodu, v okamžiku, kdy se „najednou ukáže“, že je třeba důsledně smazat vybraná data. Efektivním řešením totiž může být již od počátku všechna citlivá data už u zdroje šifrovat (ne, není to samozřejmost), takže i když pak nějaká jejich část zůstane "někde v systému", jejich obnovení je v praxi nemožné. Tím nejzásadnějším problémem ale je uvědomit si, že uvedený problém vůbec existuje. Stále se totiž

výrazně více mluví o tom, že je třeba data efektivně uchovávat a chránit, než o tom, že je třeba se jich někdy také spolehlivě zbavit. I v České republice máme případy, kdy se i citlivá osobní data vyskytovala v nechráněné podobě na místech, kde by rozhodně být neměla. A netřeba dodávat, že nejde jen o data v digitální podobě.

Bezpečné ukládání dokumentů a dat podle legislativy

Při řešení bezpečné archivace se pozornost zpravidla soustředí na záležitosti jako zálohování, dostupnost či ochrana před nepovoleným prohlížením. Má-li však digitální archiv plně nahradit listinný, je zapotřebí zajistit pro uložené dokumenty (či uložená data) onu klíčovou vlastnost, kvůli které papíry s razítky uschováváme. A sice, že takový papír mohu ještě po letech vytáhnout, předat, někam odnést, předložit partnerovi či kontrole nebo ukázat před soudem a bude jasné, že se jedná o původní dokument, do kterého nebylo nic dodatečně

vepsáno. V řadě případů je tomu tak, že teprve, je-li zajištěna tato důvěryhodnost, má smysl investovat do zálohování, dostupnosti atd.

Papírové dokumenty mají řadu nevýhod (nelze je zálohovat ani pořádně řídit přístupy, vyhledávání je pracné atd.), jejich výhodou však je poměrně jednoznačná shoda o tom, co bude protistrana akceptovat jako doklad důvěryhodnosti. Podpis, razítko, případně notářské ověření. U digitálních dokumentů a dat se zdá, že situace je komplikovanější vývojem technologií, které mohou vypadat zcela jinak než dnes a s nimi se mohou měnit i požadavky na prokazování pravosti. Je proto velkou výhodou, že existuje jednoznačný právní předpis, který říká, jak mají být data a dokumenty ošetřeny, aby byly uznány za pravé, a to i v jakkoliv vzdálené budoucnosti.

Tímto předpisem jsou specifikace vydané Evropským institutem pro telekomunikační standardy (ETSI), které naprosto detailně popisují, jaká ověřovací metadata má dokument obsahovat a jak má být o něj pečováno, aby mohl být i v budoucnu uznán za pravý „nade vší pochybnost.“ Na tyto

standardy odkazuje platné rozhodnutí Evropské komise, zákony členských států (včetně českého zákon o elektronickém podpisu) a odkazuje na ně i momentálně schvalovaná Regulace o elektronické identifikaci a důvěryhodných službách pro elektronické transakce na vnitřním trhu.

Kdo chce mít jistotu, že uložená data a dokumenty těmto normám vyhovují, v má různé možnosti. Může studovat právníké a technické detaily norem, vyčlenit pár kolegů, investovat do vývoje a testování a rozšířit své systémy o potřebné funkce. Další možností je koupě speciálního archivačního zařízení, které bude příslušným normám vyhovovat. Zde je však důležité nespokojit se s tvrzením výrobce a požadovat jednoznačný důkaz.

A pak je tu služba Long-Term Docs. Dokumenty necháte v dosavadním úložišti (vlastním nebo cloudovém), nic neměníte a služba na dálku zařídí, že dokument bude vytvořen správným způsobem (ve shodě se všemi relevantními normami) a v budoucnu bude řádně ošetřován. Všechny úkony jsou automatizovány. Je při tom podstatné, že:

- dokument neopouští úložiště a neopouštějí jej

ani žádná metadata, ze kterých by bylo možné dovodit jeho obsah. Ani potenciálně tedy nehrozí narušení bezpečnosti;

- za průběžnou péči o dokument (včetně přidávání čerstvých časových razítek) nejsou účtovány žádné poplatky. Nehrozí tedy, že by se dlouhodobá digitální archivace prodražila.

Jak již bylo uvedeno, služba Long-Term Docs pokrývá dvě oblasti práce s dokumenty nebo datovými balíky.

Za prvé: Správné vytvoření. Je zřejmé, že sebelepší péče o dokument nepomůže, pokud je podpis v nesprávném formátu nebo kupříkladu chybí ověřovací metadata. V tom jsou digitální dokumenty stejné jako papírové - listině bez razítka také nepomůže sebepečlivější archivace. Služba Long-Term Docs tedy zkontroluje formát dokumentu, provádí konverzi do archivačního PDF/A (je-li to vhodné), zkontroluje přítomnost elektronického podpisu nebo jej připojí, spojí se s certifikační autoritou a získá od ní metadata potřebná pro ověření pravosti, připojí tato metadata k dokumentu a na závěr přidá časové razítko tak, aby kontrolní

otisk (hash) chránil i veškeré metadata.

Za druhé: Správnou péčí o dokument nebo balík dat. To zahrnuje hlídání platnosti časového razítka, a pokud se tato platnost blíží vypršení, připojení čerstvého při zachování principu digitální kontinuity (nové razítko musí být připojeno před vypršením platnosti staršího, včetně ověřovacích metadat). Způsob připojování čerstvých časových razítek je rovněž detailně popsán normou ETSI.

Služba Long-Term Docs pokrývá všechny tři normy pro důvěryhodné elektronické dokumenty vydané ETSI.

PAdES LTV je určena pro archivaci důvěryhodných dokumentů ve formátu PDF. Počítá s tím, že tyto dokumenty mohou obsahovat přílohy, takže „pédéefko“ se vlastně stává kontejnerem pro úschovu dalších textů, obrázků nebo jiných datových souborů. Pochopitelně s tím omezením, že zatímco u vlastního PDF je možné zajistit bezproblémové zobrazení i na budoucích zařízeních (dosáhne toho využitím archivační verze PDF/A), u některých příloh to možné není. Je kupříkladu docela možné, že pokud bude v roce 2030 zapotřebí přehrát videozáznam z roku 2014, bude to vyžadovat

speciální zařízení. Ošetření podle PAdES LTV ovšem umožní prokázat autenticitu i takových příloh. XAdES-A je určena pro archivaci XML dat. To mohou být datové zprávy, výstupy z informačních systémů nebo cokoliv jiného. Opět platí, lze zahrnout i přílohy.

CAdES-A je určena pro archivaci dokumentů nebo dat v libovolném formátu.

Lepší než dlouhé povídání je ale službu Long-Term Docs rovnou vyzkoušet. Najdete ji na www.longtermdocs.eu, kde registrací získáte bezplatné ošetření první dávky dokumentů.

(Partnerský příspěvek)